

REGRESI DAN KORELASI LINIER SEDERHANA

8

Objektif :

1. Mahasiswa Mampu Memahami Mengenai Regresi dan Korelasi Linier Sederhana
 2. Mahasiswa Mampu Mengaplikasikan Regresi dan Korelasi Linear Sederhana pada R
-

8.1 Pengertian Regresi Linear

Persamaan matematika yang memungkinkan kita meramalkan nilai-nilai suatu variabel terikat (*dependent*) dari nilai-nilai satu atau lebih variabel bebas (*independent*) disebut persamaan regresi. Dalam bab ini, kita akan membicarakan masalah pendugaan atau peramalan nilai variabel terikat Y berdasarkan variabel bebas X yang telah diketahui nilainya.

Misalkan kita ingin meramalkan nilai kimia mahasiswa tingkat persiapan berdasarkan skor tes intelegensia yang diberikan sebelum mulai kuliah. Untuk membuat pendugaan semacam ini, pertama-tama kita perhatikan sebaran nilai kimia untuk berbagai skor tes intelegensia yang dicapai oleh mahasiswa-mahasiswa tahun sebelumnya. Dengan melambangkan nilai kimia seseorang dengan y dan skor tes intelegensianya dengan x , maka data setiap anggota populasi dapat dinyatakan dalam koordinat (x, y) . Data tersebut kemudian diplotkan untuk melihat apakah kedua variabel tersebut saling berhubungan secara

linear. Bila hubungan linear ini ada, maka kita berusaha menyatakan secara matematik dengan sebuah persamaan garis-lurus yang disebut garis regresi linear dalam bentuk

$$\hat{y} = a + bx$$

Dalam hal ini, a menyatakan intersep atau perpotongan dengan sumbu tegak, dan b adalah kemiringan atau gradiennya. Lambang \hat{y} digunakan di sini untuk membedakan antara nilai ramalan yang dihasilkan garis regresi dan nilai pengamatan y yang sesungguhnya.

8.1.1 Pendugaan Parameter

Setelah kita memutuskan akan menggunakan persamaan regresi linear, maka kita menghadapi masalah bagaimana memperoleh rumus untuk menentukan nilai dugaan titik bagi a dan b berdasarkan data sampel. Untuk itu, kita menggunakan prosedur yang disebut metode kuadrat terkecil (*least square method*), yaitu metode yang memilih suatu garis regresi yang membuat jumlah kuadrat jarak vertical dari titik-titik pengamatan ke garis regresi (*error*) tersebut sekecil mungkin.

Bila diberikan data contoh $\{(x_i, y_i); i = 1, 2, \dots, n\}$ maka nilai dugaan kuadrat terkecil bagi parameter dalam garis regresi

$$\hat{y} = a + bx$$

Dapat diperoleh dari rumus

$$b = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

dan

$$a = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n}$$

Contoh Kasus:

Berikut ini adalah pengaruh kualitas layanan terhadap kepuasan nasabah di Bank ABC ditunjukkan dalam tabel berikut ini :

| Kualitas Layanan | Kepuasan Nasabah |
|------------------|------------------|
| 9 | 58 |
| 8 | 89 |
| 5 | 80 |
| 1 | 55 |

Tentukan persamaan regresinya dan dugalah kepuasan nasabah apabila kualitas layanan yang diberikan sebesar 4.

Penyelesaian :

Karena kualitas layanan yang mempengaruhi kepuasan nasabah, maka variabel bebas X pada kasus ini adalah Kualitas Layanan, sedangkan Kepuasan Nasabah merupakan variabel terikat Y . Kita peroleh bahwa

$$\sum X = 23 \quad \sum X^2 = 171 \quad (\sum X)^2 = 529 \quad \sum XY = 1.689$$

$$\sum Y = 282 \quad \sum Y^2 = 20.710 \quad (\sum Y)^2 = 79.524 \quad n = 4$$

Sehingga,

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2}$$

$$b = \frac{4(1689) - (23)(282)}{4(171) - (529)}$$

$$b = \frac{6756 - 6486}{684 - 529} = 1,7419 \approx 1,742$$

$$a = \frac{\sum Y - b\sum X}{n}$$

$$a = \frac{282 - 1,7419(23)}{4}$$

$$a = \frac{241,9363}{4} = 60,484075 \approx 60,484$$

Dengan demikian, persamaan regresinya adalah:

$$\hat{y} = 60,484 + 1,742 x$$

Kita dapat menduga kepuasan nasabah apabila kualitas layanan senilai 4 dengan mensubstitusikan nilai $x = 4$ ke dalam persamaan, sehingga

$$\hat{y} = 60,484 + 1,742 x = 60,484 + 1,742 * 4 = 67,452$$

Artinya, kita menduga kepuasan nasabah akan bernilai $67,452 \approx 67$ apabila kualitas layanannya senilai 4.

8.1.2 Korelasi Linier

Dalam subbab ini, kita membicarakan masalah pengukuran hubungan antara dua variabel X dan Y , dan bukan meramalkan nilai Y dari pengetahuan mengenai variabel bebas X seperti dalam regresi linear. Pada analisis korelasi, kita mencoba mengukur kekuatan hubungan antara dua variabel tersebut melalui sebuah bilangan yang disebut **koefisien korelasi**.

Kita mendefinisikan koefisien korelasi linear sebagai ukuran hubungan linear antara dua variabel X dan Y , dan dilambangkan dengan r . Ukuran korelasi linear ini yang paling banyak digunakan adalah koefisien korelasi **Pearson** dengan rumus

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n(\sum X^2) - (\sum X)^2][n(\sum Y^2) - (\sum Y)^2]}}$$

Keterangan :

1. Jika $r = 0$, maka tidak ada hubungan linear antara kedua variabel
2. Jika $r = (-1)$, maka hubungan sangat kuat dan bersifat tidak searah
3. Jika $r = (+1)$, maka hubungan sangat kuat dan bersifat searah

jadi, hubungan linear sempurna terjadi bila $r = +1$ atau -1 . Bila r mendekati $+1$ atau -1 , hubungan antara kedua variabel tersebut kuat dan kita katakan terdapat korelasi yang tinggi antara keduanya. Akan tetapi, bila r mendekati nol, hubungan linear antara kedua variabel sangat lemah atau mungkin tidak ada sama sekali.

Kesalahan apabila disimpulkan bahwa $r = 0,6$ menunjukkan adanya hubungan linear yang dua kali lebih kuat daripada yang ditunjukkan oleh $r = 0,3$. Di pihak lain, apabila besarnya r^2 , yang biasanya disebut sebagai **koefisien determinasi**, maka kita mempunyai bilangan yang menyatakan **proporsi keragaman total nilai-nilai variabel Y yang dapat dijelaskan oleh nilai-nilai variabel X melalui hubungan linear tersebut**. Jadi, suatu korelasi sebesar $r = 0,6$ bermakna bahwa 0,36 atau 36% di antara keragaman total nilai-nilai Y dalam sampel kita dapat dijelaskan oleh hubungannya dengan nilai-nilai X .

Contoh Kasus :

Berikut ini adalah pengaruh kualitas layanan terhadap kepuasan nasabah di Bank ABC ditunjukkan dalam tabel berikut ini :

| Kualitas Layanan | Kepuasan Nasabah |
|------------------|------------------|
| 9 | 58 |
| 8 | 89 |

| | |
|---|----|
| 5 | 80 |
| 1 | 55 |

Tentukanlah koefisien korelasi dan koefisien determinasi

Penyelesaian :

Diketahui :

$$\sum X = 23 \quad \sum X^2 = 171 \quad (\sum X)^2 = 529 \quad \sum XY = 1.689$$

$$\sum Y = 282 \quad \sum Y^2 = 20.710 \quad (\sum Y)^2 = 79.524 \quad n = 4$$

Koefisien korelasi (r) :

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n(\sum X^2) - (\sum X)^2][n(\sum Y^2) - (\sum Y)^2]}}$$

$$r = \frac{4(1689) - (23)(282)}{\sqrt{[4(171) - (529)][4(20710) - (79524)]}}$$

$$r = 0,3766$$

Koefisien determinasi (r^2) :

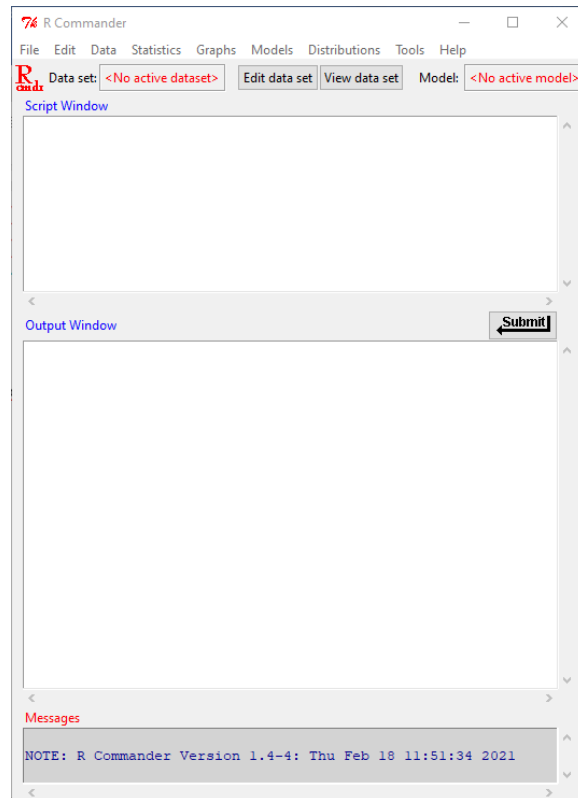
$$r^2 = 0,3766^2$$

$$r^2 = 0,1418 \rightarrow (14,18\%)$$

koefisien korelasi sebesar 0,3766 menunjukkan adanya hubungan linear yang cukup rendah antara X dan Y. karena $r^2 = 0,1418$ maka kita dapat mengatakan bahwa 14,18% di antara keragaman dalam nilai-nilai Y dapat dijelaskan oleh hubungan linearnya dengan X.

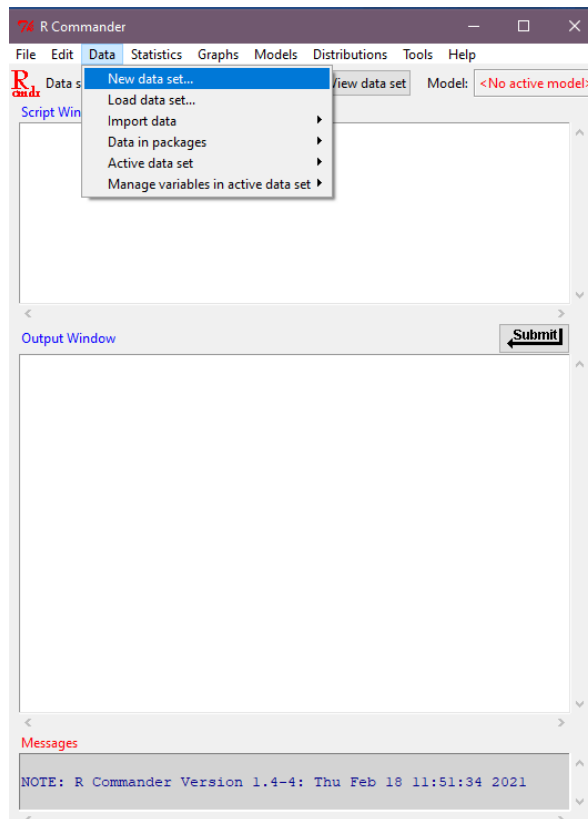
LANGKAH-LANGKAH Pengerjaan Software

1. Buka aplikasi R-commander, maka akan muncul tampilan seperti pada gambar di bawah ini :



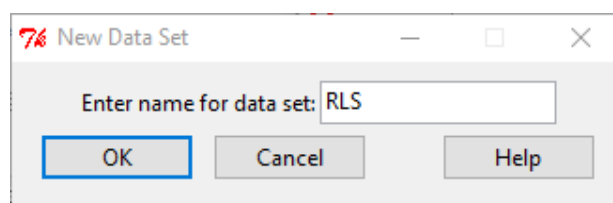
Gambar 8.2 Tampilan Awal R-Commander

2. Klik Data, lalu klik *New Data Set*, seperti pada gambar di bawah ini :



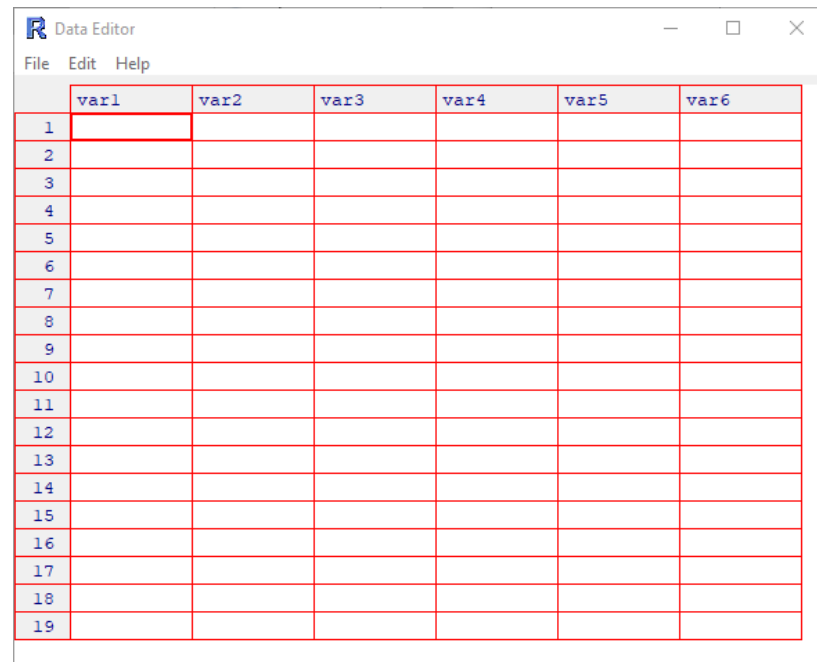
Gambar 8.3 Tampilan Langkah *New Data Set*

3. Lalu akan muncul tampilan *New Data Set*, ketika 'RLS', seperti pada gambar di bawah ini.

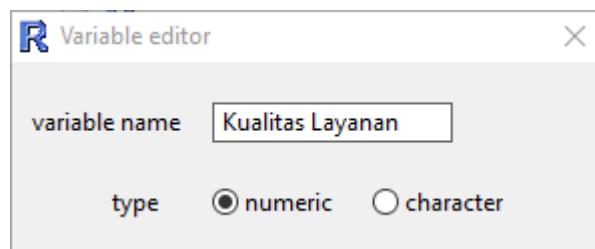
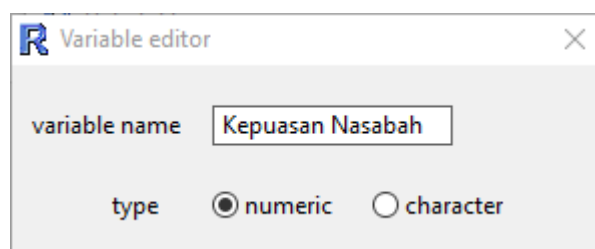


Gambar 8.4 Tampilan *New Data Set*

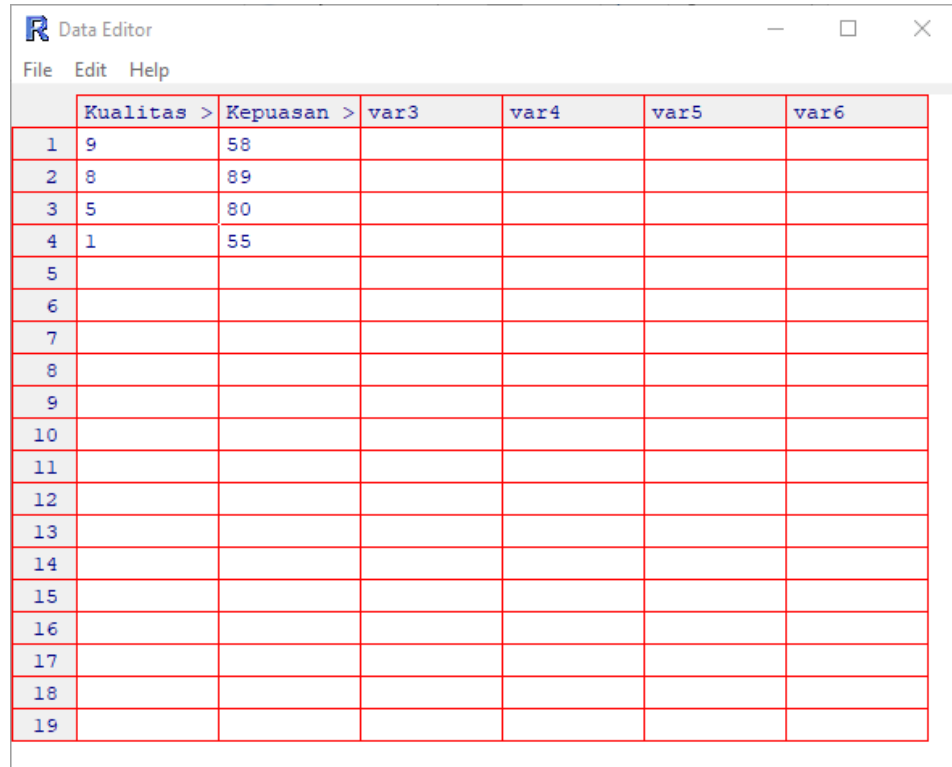
4. Klik OK, kemudian akan muncul data editor seperti di bawah ini.

**Gambar 8.5 Tampilan *Data Editor***

5. Klik var1 lalu ganti menjadi 'Kualitas Layanan', pada pilihan *type* pilih *numeric*, lalu klik var2 kemudian ganti menjadi 'Kepuasan Nasabah', pada pilihan *type* pilih *numeric* seperti gambar di bawah ini.

**Gambar 8.6 Tampilan var1****Gambar 8.7 Tampilan var2**

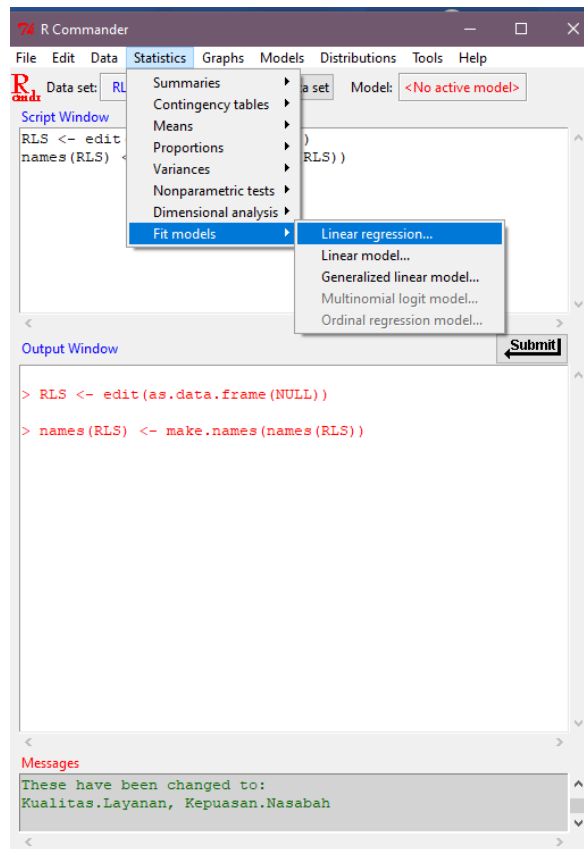
6. Kemudian isi setiap kolom sesuai dengan contoh yang diketahui seperti pada gambar di bawah ini, setelah itu klik *Close tab Data Editor*.



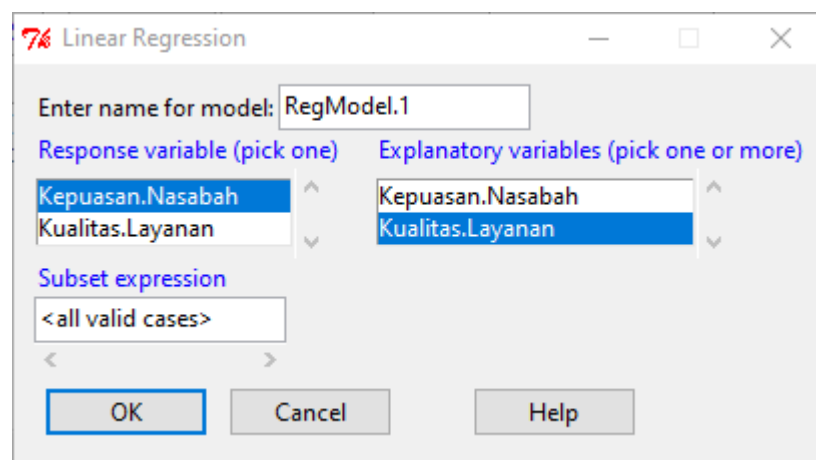
| | Kualitas > | Kepuasan > | var3 | var4 | var5 | var6 |
|----|------------|------------|------|------|------|------|
| 1 | 9 | 58 | | | | |
| 2 | 8 | 89 | | | | |
| 3 | 5 | 80 | | | | |
| 4 | 1 | 55 | | | | |
| 5 | | | | | | |
| 6 | | | | | | |
| 7 | | | | | | |
| 8 | | | | | | |
| 9 | | | | | | |
| 10 | | | | | | |
| 11 | | | | | | |
| 12 | | | | | | |
| 13 | | | | | | |
| 14 | | | | | | |
| 15 | | | | | | |
| 16 | | | | | | |
| 17 | | | | | | |
| 18 | | | | | | |
| 19 | | | | | | |

Gambar 8.8 Tampilan Data Editor yang telah diisi

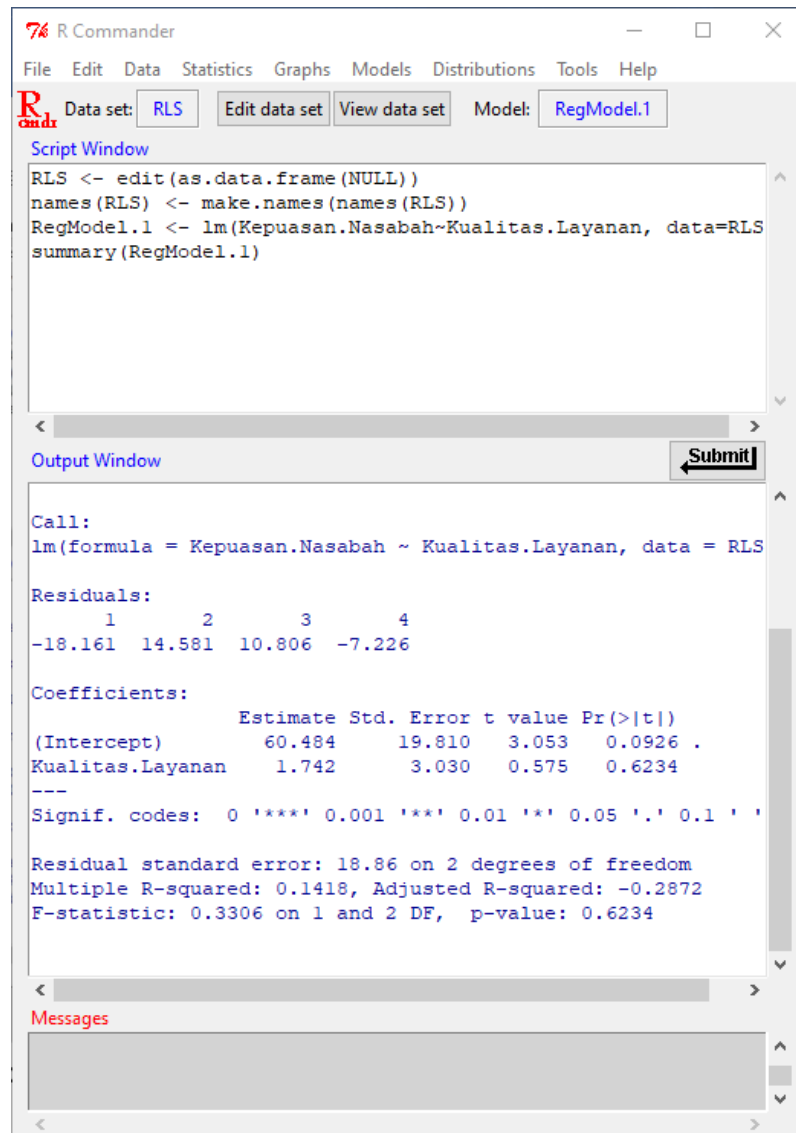
7. Kemudian klik *Statistics*, pilih *Fit models*, lalu pilih *Linier regression* seperti pada gambar di bawah ini.

Gambar 8.9 Tampilan Langkah *Linier Regression*

8. Pilih 'Kepuasan Nasabah' sebagai variabel terikat pada *Response variabel* dan pilih 'Kualitas Layanan' sebagai variabel bebas pada *Explanatory variabels* seperti pada gambar di bawah ini.

Gambar 8.10 Tampilan Box *Linier Regression*

9. Kemudian klik OK, maka akan muncul hasil output seperti pada gambar di bawah ini.



Gambar 8.11 Tampilan Output Software

```

Call:
lm(formula = Kepuasan.Nasabah ~ Kualitas.Pelayanan, data = RLS)

Residuals:
    9      8      5      1 
-18.161  14.581  10.806  -7.226 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    60.484      19.810   3.053  0.0926 .
Kualitas.Pelayanan  1.742       3.030   0.575  0.6234
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.86 on 2 degrees of freedom
Multiple R-squared:  0.1418, Adjusted R-squared:  -0.2872
F-statistic: 0.3306 on 1 and 2 DF, p-value: 0.6234

```

Sehingga diperoleh persamaan regresinya adalah $\hat{y} = 60,484 + 1,742 x$ dengan koefisien determinasi $r^2 = 14,18\%$ yang hasilnya sama dengan perhitungan manual.

Referensi :

- [1] Lind, Douglas, William G. Marchal, Samuel A. Wathen. 2006. *Basic Statistics for Bussiness and Economics* (5th edition). New York: The McGraw-Hill Companies.
- [2] Walpole, R. 1995. Pengantar Statistika. Jakarta: PT Gramedia